Episodic versus Semantic Memory: An Exploration of Models of Memory Decay in the Serial Attention Paradigm

Chris R. Sims (simsc@rpi.edu)
Cognitive Science Department
Rensselaer Polytechnic Institute

Wayne D. Gray (grayw@rpi.edu)
Cognitive Science Department
Rensselaer Polytechnic Institute

Abstract

We investigate the implications of an episodic versus semantic memory representation in a model of serial task switching. Altmann and Gray (2002; 2004) have developed a model in which the rate of memory activation, decay, and interference are tightly coupled in an explanation of learning, forgetting, and performance. A central assumption of their model is the use of an episodic memory trace for task cues. The current research explores the centrality of this assumption by exploring the predictions made by three mathematical formulations of interference and decay.

Introduction

It is clear that we can have memories for individual events such as the birth of a child or buying our first new car. It is also clear that we have memories for common facts such as the name of the President of the United States or the number of states in the Union. Tulving (1972) advocated a distinction between these types of memory, calling the former *episodic* and the latter *semantic* memories. Typically, episodic memory is viewed as an instance-based representation: each perceptual experience results in the encoding of a new instance of memory. In contrast, semantic memory is typically viewed as a non-instance based representation. Although we have all encountered the fact that 2+2=4 hundreds of times in our lives, we might only have one memory representation of this fact.[1] Although memory and memory representation are popular topics to study, it is rare to find a phenomenon in which the rate of memory activation, decay, and interference are tightly coupled in an explanation of learning, forgetting, and performance. It is even rarer to find a phenomenon where both the speed of performance and number of errors are tightly predicted. It is rarer still to find a precise mathematical formulation of such a phenomenon that depends, as its most basic assumption, on the idea that the memory it is studying is episodic, not semantic.

The Altmann and Gray (2002, 2004) theory of *serial attention* provides this rare combination of constraints. In the next section we introduce the paradigm they used and their results, as well as the *functional decay* model

they propose to account for these findings. In the following section we examine three alterations to the original model. The first holds all but the assumption of episodic memory constant, while the second two compare the predictions of two different equations of activation and decay. We intend our discussion to shed light on the functional differences between episodic and semantic memory representations and the implications of these differences for cognitive modeling. We also believe that our close comparison of three popular mathematical formulations of interference and decay sheds light on the strengths and weaknesses of each. Other modelers who depend on one or more of these formulations should find our comparisons to be at least interesting and, perhaps, disturbing.

Task Switching and the Serial Attention Paradigm

The study of task switching examines the costs and mechanisms required to rapidly switch from one task to another. In the basic task switching experiment, participants either perform one task repeatedly or switch between different tasks. In some experiments (e.g., Logan & Bundesen, 2003) each trial is preceded by a task cue, indicating which task to perform, whereas other experiments use task cues that are implicit in the trial stimuli themselves (e.g., Allport & Wylie, 2000). In general, participants who are required to switch tasks perform more slowly as compared to those performing consecutive blocks of the same task, a phenomenon generally labeled *switch cost*. Allport, Styles, and Hseih (1994) showed that the magnitude of switch cost can be reduced, but not eliminated, by increasing the interval or rest period between tasks. A variety of related phenomena have also been discovered, including restart cost, residual switch cost, and within-run slowing (Altmann & Gray, 2002).

Unfortunately, the wealth of data available in this research domain does not converge on an explanation. The earliest findings in this field were reported by Jersild (1927), and since that time dozens of different theories have been proposed. The explanations range from postulating executive control mechanisms responsible for reconfiguring cognition and selecting tasks (e.g., Monsell, 2003; Rogers & Monsell, 1995; Kieras, Meyer, Ballas & Lauber, 2000), to attempting to account for task switching phenomena as a result of some combination of well-

---

[1] A notable exception to this classification is Logan's (2002) instance theory of memory, in which both semantic and episodic memories are represented as instances.
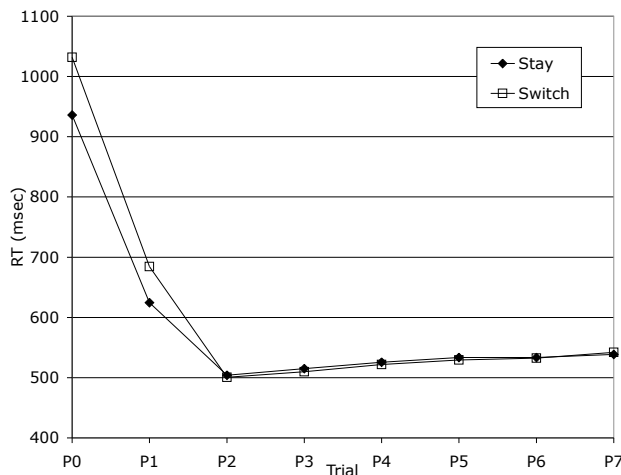
Figure 1: Mean response time by task continuity. Data taken from Altmann and Gray (2004).

established mechanisms (e.g., Altmann, 2003; Altmann & Gray, 2004). (A more classic way of phrasing the difference is as between explanations of task switching that postulate new *hypothetical constructs* versus those constructed out of established *intervening variables*, see MacCorquodale & Meehl, 1948).

Altmann and Gray (2004) favor the intervening construct approach and have constructed a model of task switching that requires no special executive control processes. Instead, their model predicts complex behavior almost entirely from basic memory processes such as interference and decay. The next section will provide an overview of their paradigm and experimental results. Following this, their functional decay model and its underlying assumptions will be examined in detail.

Paradigm and Results

In their experiment 1A, Altmann and Gray (2004) presented participants with one of two tasks on each trial. The two tasks were a Groupsize task or a Position task for a string of repeated letters (for example, 'AAAA'). The Groupsize task was to determine whether a string contained greater or fewer than five characters, while the Position task was to determine whether the letters occurred at the beginning or end of the alphabet. The task trials were grouped into runs, such that participants would perform one task for the duration of the run. Since the trial stimuli were ambiguous as to task, before the start of each run an instructional cue would indicate whether the Groupsize or the Position task was to be performed. This cue would remain on the screen until dismissed by a keypress. Run lengths were chosen randomly according to a uniform distribution in the range $7 \leq n \leq 13$, constrained so that each successive pair of runs would sum to 20 trials. Each successive pair defined a task block.

Figure 1 shows the mean response latency for the second run of each block, separated by continuity; that is, whether the task *switches* or *stays* the same. P0 repre-

sents the presentation of the task cue, while P1 through P7 are the first seven trials of that task in the run. This figure demonstrates several phenomena relevant to task switching. First, switch cost is observed as the difference between P0 Switch and P0 Stay. The data also shows residual switch cost, as the lingering latency difference into the first trial of the run. Finally, and most relevant to this research, within-run slowing is observed as the gradual but steady decline in performance over trials 2 through 7. In their experiment 3, Altmann and Gray varied the task runlength between participants. Using runlengths of 10 and 20, they found that the slope of within-run slowing is sensitive to the runlength of the task, such that longer runlengths result in a more gradual decay than shorter runlengths.

Functional Decay Model

To explain these various phenomena, Altmann and Gray assume that a task cue for the current task is retained in memory. On each trial, participants must successfully retrieve this cue in order to know which of the two tasks to perform. Additionally, Altmann and Gray assume an episodic, instance-based representation of task cues. Each time a task cue is encountered at the beginning of a run, it is encoded in memory as a new trace, regardless of whether the current cue was identical to the task cue from the previous run. According to their functional decay model, this process of encoding corresponds to rapidly rehearsing the task cue until it reaches a threshold level of activation.

In the empirical data in Figure 1, the process of encoding can account for the relatively slow response time for P0 as compared to P2 through P7. Residual encoding is also used to explain the lingering latency on trial P1. The functional decay model predicts that the activation of a task cue decreases from its peak encoding value. The reasoning behind this is that if task cues did not decay in memory, each successive encoding would have to reach a higher threshold than the previous, resulting in an escalating spiral of encoding. The theory that task cues decay across each run also explains the within-run slowing phenomenon: in the memory model used (discussed in the next section) retrieval time is a function of activation level. Hence, as declarative memory chunks decay, their retrieval times increase as well as becoming more prone to retrieval failures. Thus, across a run both retrieval time and errors should increase. Using this functional decay model, Altmann and Gray were able to obtain close quantitative fits to their data.

Modeling Memory

Each of the three memory functions considered in this paper share the three main assumptions of functional decay theory. First, each assumes an activation account of memory. This account proposes that the encoding of information imparts the memory representation (or trace) with an activation that decays as a function of time. Second, each assumes that interference among memory elements may affect the success of retrieving the correct memory. Third, a key feature of the functional de-

cay model is the notion that as memories decay they become more susceptible to the effects of interference. Although decay and interference have historically been seen as competing accounts of memory, in recent years the use of decay in combination with interference has been justified by a large and growing body of research (see Altmann & Schunn, 2002 for a review).

According to the ACT-R cognitive architecture (Anderson & Lebiere, 1998), memory is stored in declarative and procedural memory subsystems. For the purpose of this paper, declarative memory is the system of interest, as it encompasses both episodic and semantic memory. Knowledge in declarative memory is represented in the form of memory chunks, each of which has a base level of activation that decays according to a power law of forgetting, and increases through rehearsal. Equation 1 shows the ACT-R chunk activation function.

$$a_i = ln(\sum_{j=1}^{n} t_j^{-d}) \qquad (1)$$

In this function, a_i is the activation of chunk i, t_j is the time lapse since its jth retrieval, and n is the total number of retrievals during the lifetime of the chunk. d is a parameter governing the rate of decay, typically set to the value 0.5. Assuming that retrievals are evenly spaced across the lifetime of the chunk, the summation in Equation 1 can be replaced by $N \cdot (N/T)^{-d}$, where N is the total number of retrievals, and T is the total lifespan of the chunk. This then reduces to the function given by Equation 2:

$$a_i = ln(\frac{N}{1-d}) - d \cdot ln(T) \qquad (2)$$

For computational efficiency reasons, Equation 2 is used by default in the ACT-R implementation, and can be set with the ACT-R :ol parameter. However, as will be discussed, the choice of episodic or semantic memory representation has important implications for the justification of using this simplified equation.

Although other classes of equations have been considered (see Wickens, 1998, for a review), power functions have been successfully used to fit a large body of data (Anderson, 2000). Additionally, Anderson & Schooler (1991) show that power functions also find theoretical interpretation as a reflection of the probability of needing information in the natural environment. They argue that it is rational for memory to mirror the need structure of the environment.

In the subsections that follow, we consider each of the two ACT-R decay equations presented above, along with an equation proposed by Pavlik and Anderson (2003) to account for the spacing effect. Each candidate equation will be considered in the context of episodic versus semantic memory representation in a task switching paradigm.

Episodic Assumptions and the Model Used by Altmann & Gray

Altmann and Gray claim that the use of an episodic memory representation in their model is critical to the
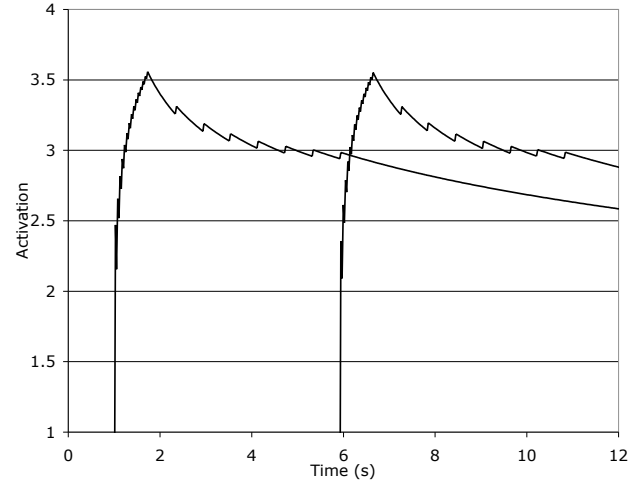

Figure 2: Memory activation trace for two task cues encoded in episodic memory.

flexibility of the cognitive system. In this section we question the necessity of this assumption in accounting for the data that their theory currently explains. In addition, we raise the issue as to whether an acknowledged limitation of the functional decay model is inherent in the model or due to their choice of an episodic rather than semantic representation.

Figure 2 shows the activation trace of two episodic task cues for the simplified decay equation used by Altmann and Gray (Equation 2). The initial spike for each task cue represents a period of rapid rehearsal while the task cue is present on the screen. Following this, the task cues are retrieved from memory once per trial. The spacing of these retrievals is insufficient to maintain the peak activation at the time of task encoding, resulting in decay. After a run of seven trials, a new task cue is presented and encoded in episodic memory, regardless of whether the same task is being performed.

In contrast to the episodic assumption, if the same two task cues are reused across runs (a non-instance based, or semantic representation), then activation becomes relatively stable as the number of retrievals increases. Hence, the activation for the memory of the current task will not differ significantly from the activation of the prior task. However, this argument is based on ACT-R's simplified decay equation. While activation does become stable at large values of N and T (see Equation 2), the equation assumes an approximately uniform spacing between retrievals. This assumption does not hold in the case of the functional decay model, since most retrievals of a trace occur in the rapid period of encoding at the beginning of the episodic trace's lifetime. Further, when task switches occur, the task cue for the previous task will not be retrieved for relatively long stretches of time.

In contrast, the complete ACT-R decay equation takes both the number and distribution in time of individual retrievals into account. As we discuss in the next section, using this form of the decay function (see Equation 1), the activation of a single (i.e., non-episodic) trace

may fluctuate enough to produce the differences in activation that are key to Altmann and Gray's theory of serial attention.

An acknowledged limit of the functional decay model is its inability to account for the variable slope of within-run slowing. In an experiment in which subjects had either an average run length of 10 or 20 trials, Altmann and Gray found that increasing the run length had the effect of decreasing the rate of slowing by approximately half. Crucially, this between-Ss difference in slowing rate was apparent across all trials of a given run. Altmann and Gray (2004) leave this effect unexplained. Indeed, it would be hard to explain in a purely episodic task cue representation without introducing additional mechanisms.

To account for variable rate within-run slowing, the cognitive system must be differentially modified by short and long runs. This possibility requires that an additional mechanism must persist *across runs* in order for it to be modified. Episodic memory chunks are unlikely to fit this description, as they are never retrieved (except in cases of retrieval error) beyond the duration of one task run. The next section explores the possibility that semantic task cues might serve this purpose in the cognitive system.

Non-episodic Functional Decay

Altmann and Gray's central argument for an episodic memory representation is the cognitive inflexibility of the decay equation over a large number of retrievals. As discussed above, this problem might be more indicative of the unsuitability of the simplified decay equation rather than a theoretical limitation of the architecture. In order to examine this possibility, we modified our MATLAB simulation to use the complete ACT-R decay equation. By conducting the simulation in MATLAB we were able to plot activation values as a continuous function of time, a task that is significantly harder in the ACT-R implementation. In addition, we modified the simulation to re-use old task cues rather than creating new memory chunks at the start of each run.

Figure 3 shows a comparison between the simplified and full ACT-R decay equations using a semantic memory representation of task cues. The sequence in Figure 3 corresponds to a participant performing the same task for four runs in a row (the *stay* condition). At the presentation of each task cue, the rehearsal is performed on an existing memory chunk rather than a newly created chunk. As predicted, the simplified decay equation quickly stabilizes on a value.

The differential effect of the two equations on predicted activation is notable even in the first run. The full curve decays faster and is more sensitive to within-run retrievals (as shown by the sharp jags) than is the simplified curve. Across runs the simplified curve decays less and less making it clear that an episodic representation is needed to support Altmann and Gray's theory of serial attention. However, it is also clear from Figure 3 that the necessity of the episodic assumption may be an artifact of the simplified equation.
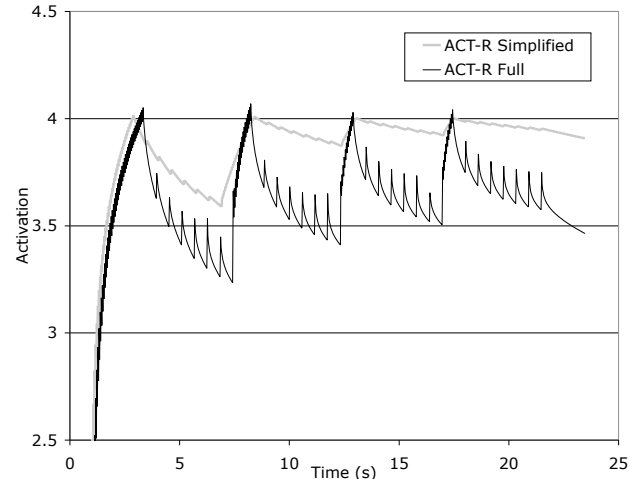

Figure 3: Comparison of the full and simplified ACT-R decay equations in a semantic memory representation.

It may be objected that although the full ACT-R equation demonstrates more flexibility with a semantic trace than does the simplified version, Figure 3 also shows that with each run of trials the activation fluctuates less with each retrieval. However, participants in the serial attention paradigm were not presented with a single task for a large number of consecutive runs. Hence, in this paradigm, the task cues would have an opportunity to decay while the second task intervenes. As far as we can determine, our modified model that assumes the full equation and semantic representation is theoretically consistent with all of the phenomena accounted for by the current functional decay model. Since task cues are re-used, on stay runs the task cue chunk starts from a higher activation. This results in less need for encoding, and therefore a P0 response time difference between switch and stay runs. Like the functional decay model, within-run slowing can be explained by the decay across runs and consequent increases in retrieval time as well as in retrieval failures.

It remains to be seen whether a semantic memory representation solves the problem of variable rate within-run slowing. Since task cue chunks now serve a functional role across runs, it seems both plausible and parsimonious for these chunks themselves to implicitly encode information about the length of a run. This model would eliminate the need for an additional cognitive mechanism that monitors the duration of each task. The question then, is how are chunks differentially modified as a consequence of long versus short run lengths?

An obvious answer is that the chunk activation itself somehow encodes information about the duration of the task. However, this alone cannot be the answer. As Figure 3 demonstrates, the primary determinant of chunk activation is not the quantity, but rather the spacing of memory retrievals. In the serial attention paradigm, trials progressed at the same speed regardless of run length. Because of this, by the end of a run, chunks that are retrieved 10 times following encoding will tend to have

similar activation levels to chunks retrieved 20 times, and so activation level alone cannot be used as a predictor for run length. Hence, although a combination of a semantic representation and the full decay equation (Equation 1) leads to the same predictions as the functional decay model, neither model seem capable of accounting for variable rate within-run slowing.

The Pavlik & Anderson Decay Function

An alternative answer may be that activation duration, rather than amplitude, leads to the observed effect of variable within-run slowing. Since longer runs result in more spaced rehearsals of a task cue, it seems plausible that these chunks might decay more gradually, and therefore lead to a decreased rate of slowing. This theory is consistent with a widely studied phenomenon in experimental psychology, known as the spacing effect. The general finding is that increasing the interval between practice episodes has the effect of increasing the probability that the studied information can be correctly recalled at a later time. Conversely, the more closely two practice sessions occur, the more likely the practiced item is forgotten.

Pavlik and Anderson (2003) proposed a model of the spacing effect that utilizes a modified form of the decay function shown in Equation 1. This model assumes that the decay rate of a chunk is controlled not only by the frequency of retrieval, but also by the activation of the chunk at the time of its retrieval. Equations 3 and 4 give the Pavlik and Anderson decay function.

$$d_j = ce^{a_j} + b \quad (3)$$

$$a = ln(\sum_{j=1}^{n} t_j^{-d_j}) \quad (4)$$

Equation 4 is identical to the full ACT-R decay function, except the decay constant d is replaced by the function d_j. In Equation 3, a_j is the activation of chunk a at the time of its jth retrieval, while b and c are scaling parameters. These equations imply that retrievals of a chunk with a high level of activation have little impact on the long-term behavior of that chunk. At the same time, infrequent but steady retrievals will cause the memory chunk to decay more slowly.

Using Equation 4 in a non-episodic version of Altmann and Gray's functional decay model should have two consequences. First, the high density of retrievals during the rehearsal phase should have relatively little impact on the long-term activation of the chunk, since these occur at a high activation level compared to the rest of the run. Secondly, increasing the run length of a task has the effect of increasing the number of spaced retrievals of that task cue. Therefore, longer tasks should result in memory representations that decay more slowly, and consequently demonstrate a lower rate of within-run slowing.

Figure 4 shows the activation traces of two task cues overlaid. The first task uses a run length of 20 while the second task uses a run length of 10. The parameters
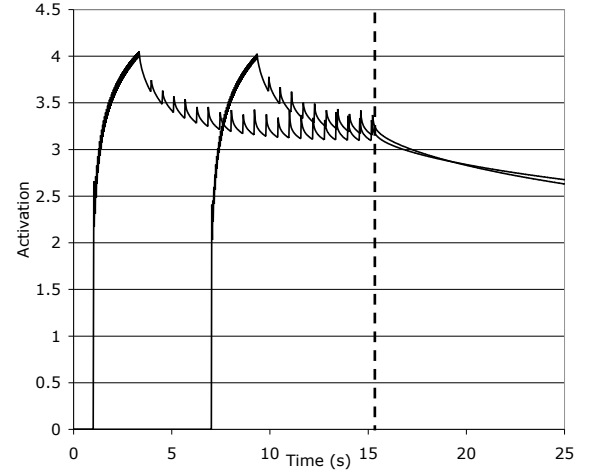

Figure 4: Comparison of decay of task cues at run lengths 20 and 10 using Pavlik & Anderson's (2003) decay equation. The dotted line indicates the disuse phase of both traces.

used in this plot were [$c = 0.232$, $b = 0.5$], although the results obtained are not sensitive to the particular values chosen. Unfortunately, the effects expected of Pavlik and Anderson's equation were not observed. The expected effect was a smaller rate of decay during disuse for the longer run-length trace. This result would have helped explain the variable rate of performance decline observed by Altmann and Gray. However, as Figure 4 shows, during the period of disuse (to the right of the dotted line), both memory traces decay at the same rate.

We hypothesize that these unexpected results are due to the high number of retrievals performed during the encoding phase of each task. Although the longer run-length task receives more spaced practice, the relative proportion of spaced practice is only slightly higher than in the shorter run-length condition. Additionally, Pavlik and Anderson's equation was developed to account for spacing differences on the order of seconds or minutes. The equation may not apply to the much shorter duration of continuous rehearsal proposed by the functional decay model.

Summary and Conclusions

This paper has examined issues regarding the use of an episodic versus semantic memory representation in cognitive modeling. It is not our intention to suggest that all memory processes must be purely semantic or purely episodic. Indeed, Tulving (1972) has suggested that any non-trivial task necessitates the use of both forms of memory. Rather, we have focused on the ways in which the choice of a memory model seems to dictate the choice of representation in the functional decay model. The exploration was grounded in the context of a well-developed model of serial task switching by Altmann and Gray (2004). A critical assumption of this model is the use of an episodic memory representation for current task cues. Although by using this approach

Altmann and Gray were able to explain a number of phenomena, some evidence suggests that their assumption is not fully justified.

By simulating the ACT-R memory activation equations we uncovered a potentially disturbing limitation of the simplified function. If memory chunks are not retrieved uniformly or at random, the decay delivered by the simplified equation will have a more shallow slope and will asymptote at a higher level than that predicted by the ACT-R theory. Since the simplified equation is used by default, many ACT-R modelers may be unaware of the consequences of equation choice. This issue is especially relevant to models using a non-instance based memory representation, as in these cases retrievals are likely to exhibit distinct patterns of use and disuse.

After simulating the full ACT-R equation, we found that contrary to Altmann and Gray, memory traces can maintain flexibility even after a relatively large number of retrievals. Building on this finding, we explored the possibility of using a semantic memory representation in the functional decay model. At a qualitative level, the semantic representation model is able to explain the same phenomena as its episodic counterpart.

We also experimented with an equation that Pavlik and Anderson (2003) developed to account for the spacing effect. By using this equation, we hoped to explain the variable rate within-run slowing phenomenon observed in Altmann and Gray's experiment 3. Although the Pavlik and Anderson equation failed to explain the intended phenomenon, its use did not diminish the performance of the general functional decay model.

Finally, it is not clear to us whether the current set of memory decay functions are applicable to the sort of rapid memory rehearsal proposed by the functional decay model. Indeed, the development and evidence for all three was in the context of data observed across timescales in the range of 10s of seconds to years. An interesting area for future research will be the investigation of how interference and decay alter behavior on the level of milliseconds.

Acknowledgments

References

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance IV* (pp. 421–452). Cambridge, MA: MIT Press.

Allport, A. & Wylie, G. (2000). Task switching, stimulus-response bindings, and negative priming. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII: Control of cognitive processes* (pp. 35-70). Cambridge, MA: MIT Press.

Altmann, E. M. (2003). Task switching and the pied homunculus: where are we being led? *Trends in Cognitive Sciences*, 7(8), 340–341.

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science*, 13(1), 27-33.

Altmann, E. M., & Gray, W. D. (2004). Beyond switch cost: Modeling memory processes in task switching. *Manuscript submitted for publication.*

Altmann, E. M., & Schunn, C. D. (2002). Integrating Decay and Interference: A New Look at an Old Interaction. *Proceedings of the 24th Annual Conference of the Cognitive Science Society.* (pp.65-70). Mahwah, NJ: Erlbaum

Anderson, J. R. (2000). *Learning and Memory* (2nd ed.). New York: Wiley

Anderson, J. R. & Lebiere, C. (1998). *The Atomic Components of Thought.* Mahwah, NJ: Erlbaum.

Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.

Jersild, A.T. (1927). Mental set and shift. *Archives of Psychology*, Whole No. 89.

Logan, G. D., (2002). An instance theory of attention and memory. *Psychological Review*, 2, 376–400.

Logan, G. D. & Bundesen. (2003) Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance*, 29(3), 575–999.

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, 55, 95–107.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134–140.

Monsell, S. (2003). Task-set reconfiguration processes do not imply a control homunculus: Reply to Altmann. *Trends in Cognitive Sciences*, 7(8), 341–342.

Pavlik Jr., P. I., & Anderson, J. R. (2003). An ACT-R model of the spacing effect. In F. Detje, D. Dorner & H. Schaub (Eds.), *Proceedings of the Fifth International Conference on Cognitive Modeling.* (pp. 177–182). Germany: Universitats-Verlag Bamberg.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2), 207–231.

Tulving, E. (1972). Episodic and semantic memory. In E.Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press

Wickens, T.D. (1998). On the form of the retention function: Comment on Rubin and Wenzel (1996): A quantitative description of retention. *Psychological Review*, 105, 379–386.